

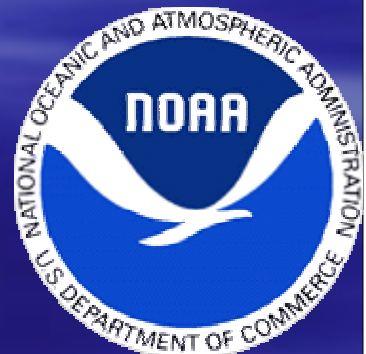
# 7<sup>th</sup> Workshop on Distributed Supercomputing – Durango, CO

---

## *NOAA Forecast Systems Laboratory Cluster*

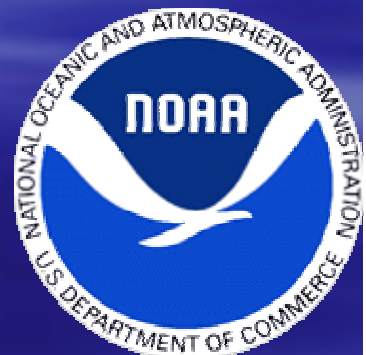
Leslie Hart  
NOAA/OAR  
Forecast Systems Laboratory  
High Performance Computing Systems  
Boulder, CO  
March 4-6, 2003

<http://www-fd.fsl.noaa.gov/hpcs>



# NOAA Organizations

- ♦ National Weather Service
  - ♦ National Centers for Environmental Prediction
    - ♦ Operational Model Production
- ♦ Oceanic and Atmospheric Research (OAR)
  - ♦ Geophysical Fluid Dynamics Laboratory
    - ♦ Climate Modeling
  - ♦ Forecast Systems Laboratory
    - ♦ Technology Transfer (Weather Forecasting)
- ♦ Satellite Operations
- ♦ Fisheries



# Current HPCS Configuration

- ◆ 140 Dual-Processor Alphas (833MHz)
- ◆ 768 Dual-Processor Xeon (2.2GHz)
- ◆ 3.337TF on LINPACK

◆ Myrinet 2000

◆ ADIC HSM

◆ I/O (0.25-1TB/day)

◆ 2-DDN SDD's (8.4 TB total)

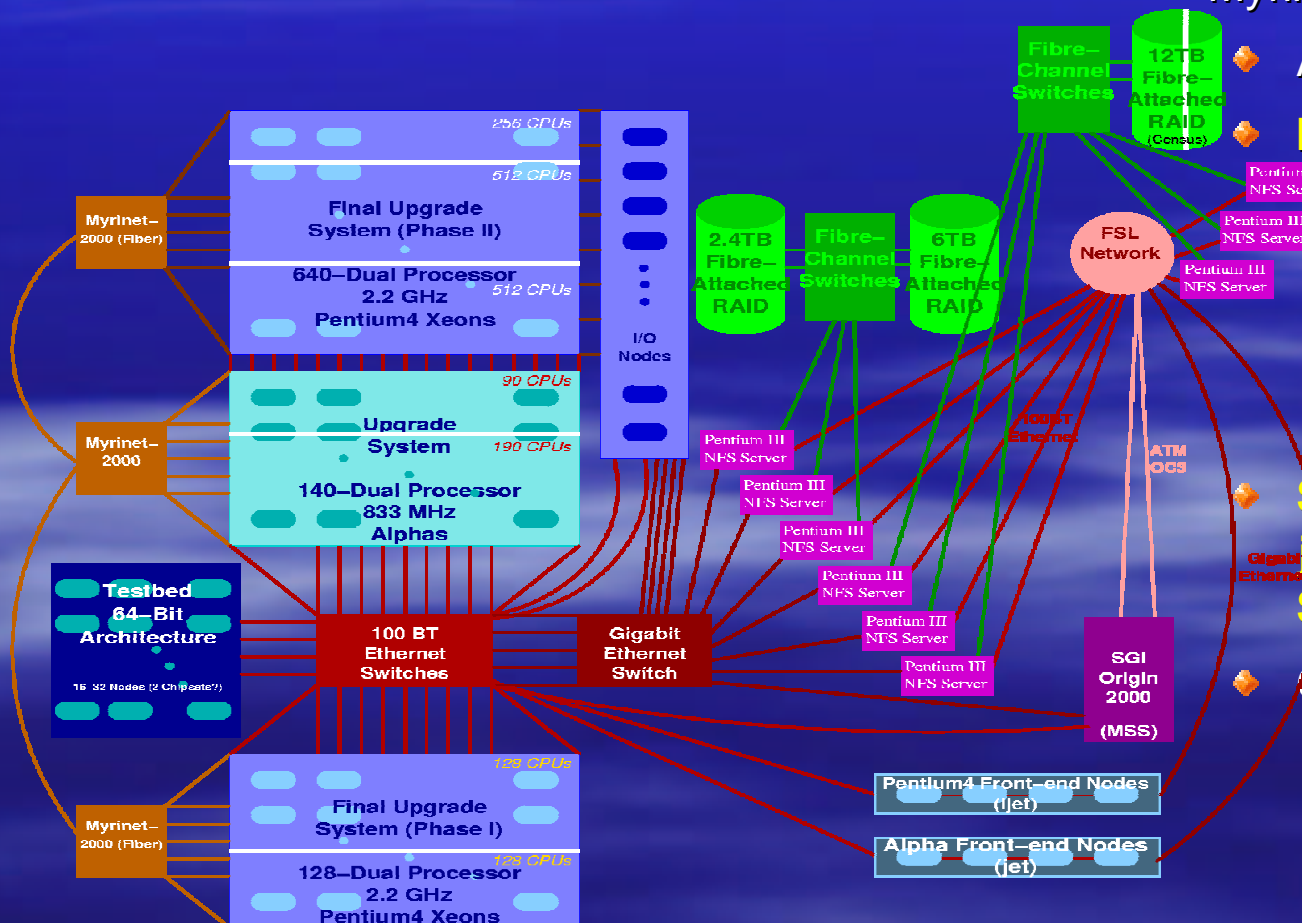
◆ Dell RAID (~12 TB total)

◆ IBM FastT700 (~2TB)

◆ NFS/XFS/CVFS/PVFS/GPFS

◆ SGE Batch System (3-5000 of jobs a day per SGE instance, 3 SGE instances)

◆ GNU/Linux OS



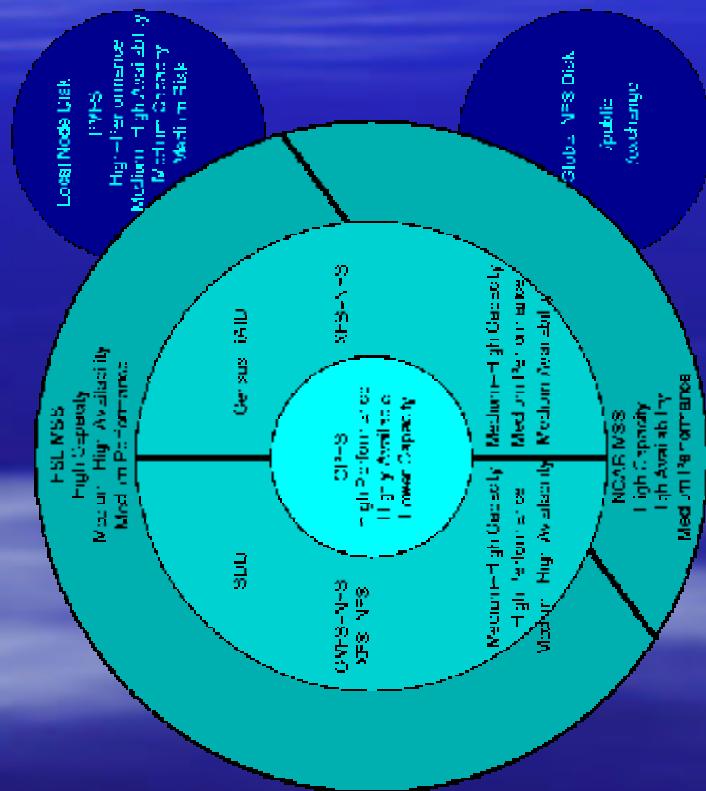
# NOAA Forecast Systems Laboratory

- ♦ Technology Evaluation and Transfer Laboratory
  - ♦ Real-time Weather Model Development
  - ♦ Weather Workstation Development
  - ♦ Observational Systems
  - ♦ Computing
- ♦ High Performance Computing Contract
  - ♦ Vendor/Integrator: HPTi
    - ♦ Subcontractor: Aspen Systems

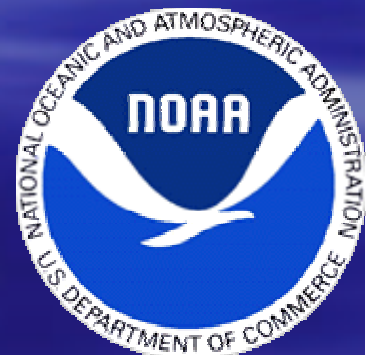




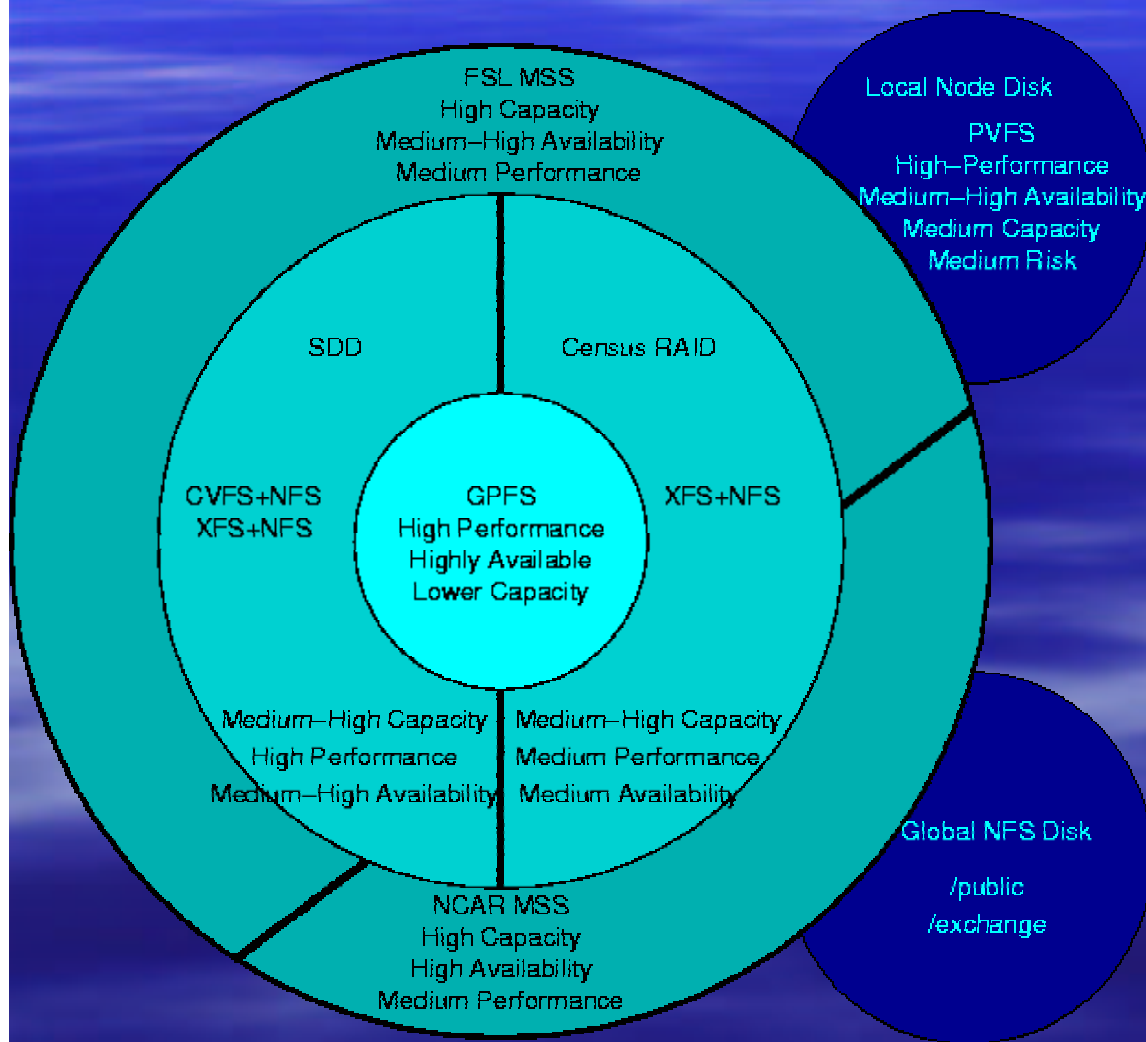
# I/O



- ◆ No satisfactory solution (yet ..., watching Lustre, trying GPFS)
- ◆ Currently using a series of NFS servers
- ◆ Using PVFS for high demand applications
- ◆ In process of trying IBM's GPFS for critical/high-impact applications



# I/O



- ◆ No satisfactory solution (yet ..., watching Lustre)
- ◆ Currently using a series of NFS servers
- ◆ Using PVFS for high demand applications
- ◆ In process of trying IBM's GPFS for critical/high-impact applications



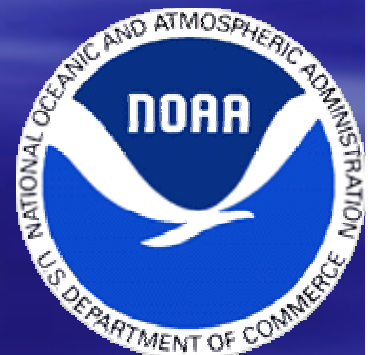
# Computing Requirements

- Current state is 20KM CONUS model
  - Requires 36 CPUs to complete in  $< 1\text{HR}$
- Short-term goal: 13KM
  - Would require  $>150$  CPUs to complete in  $< 1\text{HR}$
- Medium-term goal:  $<10\text{KM}$ 
  - Would require  $> 350$  CPUs to complete in  $< 1\text{HR}$
- Multiple copies run simultaneously
- Developmental Testbed Center
  - Run several test models at above scales



# Staff/Budget

- ♦ \$3-3.5M/year for HPTi contract
- ♦ 9 Staff members + operations staff
  - ♦ 6.5-7 FTE
    - ♦ 1 Sys Ad
    - ♦ 2 HPTi for hardware/software
    - ♦ 1 User Support
    - ♦ 1 User Liason
    - ♦ 1 Network manager
    - ♦ 2 MSS
    - ♦ 1 Technical management
  - ♦ Operations staff 24x7

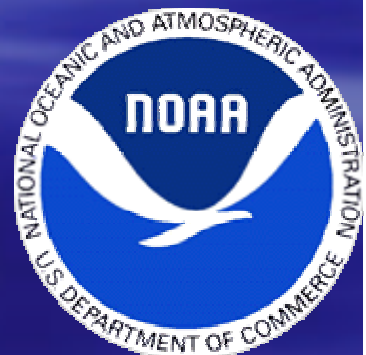




# Q1: Is the machine living up to expectations?

---

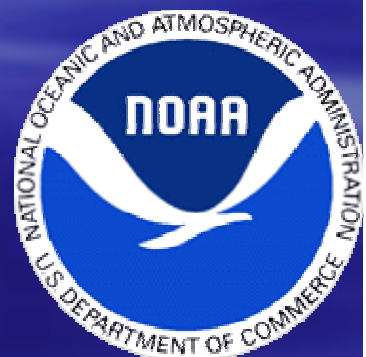
- In terms of computational performance we are seeing performance we expected (we benchmarked the actual codes). Until...we want to save the answer !
- The scaling is adequate, our codes are fairly latency tolerant.



## Q2: What is MTBI? What are the primary failure modes? What is average utilization?

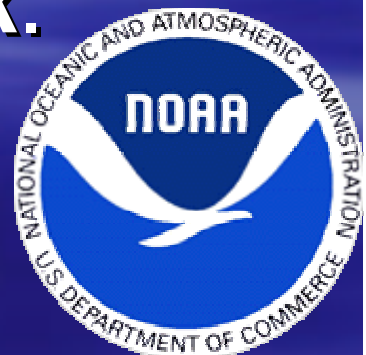
---

- Nodes fail (mostly due to software <1/week, hardware <1/month).
- Jobs fail to temporary stale NFS file handles. (Every few days.)
- Primary failures relate back to NFS servers.
- Typical utilization around 25%



# Q3: What is the primary complaint?

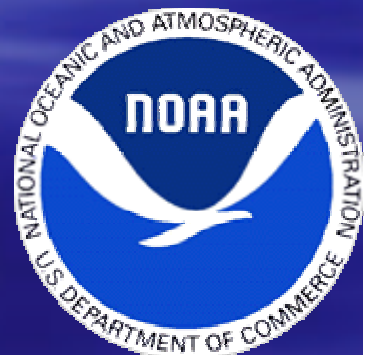
- ♦ Failure of I/O services.
  - ♦ The failures don't always manifest themselves directly as I/O failures.
- ♦ Deterministic real-time scheduling.
- ♦ Compiling problems (failure to support “common” extensions). “Things don't work exactly like they do on machine **X**.”



# Summary of Needs

---

- High-speed, highly stable, highly scalable I/O
- Memory bandwidth seems to limit performance of applications
- Software for diagnostics/monitoring
- Compilers





# HPCS Systems Activity

## File Systems Status

(Last updated 02/28 21:53 - Intel GMT)

/p10 77% Used 77 GB Used	/p11 77% Used 77 GB Used	/p12 77% Used 77 GB Used	/p13 77% Used 77 GB Used	/p20 96% Used 163 GB Used	/p21 68% Used 340 GB Used	/p22 61% Used 304 GB Used
/p30 67% Used 342 GB Used	/p31 45% Used 229 GB Used	/p32 67% Used 340 GB Used	/p40 55% Used 277 GB Used	/p41 97% Used 187 GB Used	/p42 69% Used 353 GB Used	/p43 88% Used 231 GB Used
/rt0 48% Used 79 GB Used	/rt1 39% Used 136 GB Used	/rt2 38% Used 56 GB Used	/rt3 54% Used 104 GB Used	/exchange 4% Used 38 GB Used	/public	

## Cron Servers and Front-end Status

(Last updated 02/28 21:48 GMT)

Cron0 Load Avg 0.00	Cron1 Load Avg 1.05	Cron2 Load Avg 0.04	Cron3 Load Avg 0.02	IJet-F00 Load Avg 0.23	IJet-F01 Load Avg 0.10
Jet-F00 Load Avg 0.28	Jet-F01 Load Avg 0.01	Jet-F02 Load Avg 2.44	Jet-F03 Load Avg 0.00	IJet-F02 Load Avg 2.65	IJet-F03 Load Avg 0.00

## User Activity On IJet

CPUs In Use By User				CPU Request Queued By User			
User	Number of CPUs	User	Number of CPUs	User	Number of CPUs	User	Number of CPUs
cohen	66	dobbing	32	cohen	66	harrop	66
grell	36	harrop	126	hart	198	moxley	66
hart	198	hydez	132	shin	128	tierney	66
jlee	26	lali	16				

## Compute Systems Status

**IJet System**  
1526 CPUs  
1524 CPUs Up (99%)  
1154 CPUs Used (75%)  
66 Total Jobs  
(Last updated 02/28 21:48 GMT)

**System A**  
126 CPUs  
126 CPUs Up (100%)  
0 CPUs Used (0%)  
0 Total Jobs

**System C**  
254 CPUs  
254 CPUs Up (100%)  
208 CPUs Used (81%)  
12 Total Jobs

**System D**  
510 CPUs  
510 CPUs Up (100%)  
482 CPUs Used (94%)  
14 Total Jobs

**System E**  
510 CPUs  
508 CPUs Up (99%)  
398 CPUs Used (78%)  
39 Total Jobs

**Alpha Jet System (Real-time only)**  
292 CPUs  
280 CPUs Up (95%)  
80 CPUs Used (27%)  
7 Total Jobs  
(Last updated 02/28 21:38 GMT)

**Alpha System A**  
194 CPUs  
184 CPUs Up (94%)  
80 CPUs Used (41%)  
7 Total Jobs

**Alpha System B**  
98 CPUs  
96 CPUs Up (97%)  
0 CPUs Used (0%)  
0 Total Jobs

## Color Code Key

File Systems Status	Satisfactory	Questionable	More than 80% full	Down	Planned Outage
Compute Systems Status	Satisfactory	Questionable	More than 95% full	Down	Planned Outage
			Less than 60% available	Down	Planned Outage
			Less than 90% available		